Databases on Kubernetes Using a Custom Operator

Day 1, Day 2, and Beyond

@unterstein @neo4j @KubeCon_ #kubecon

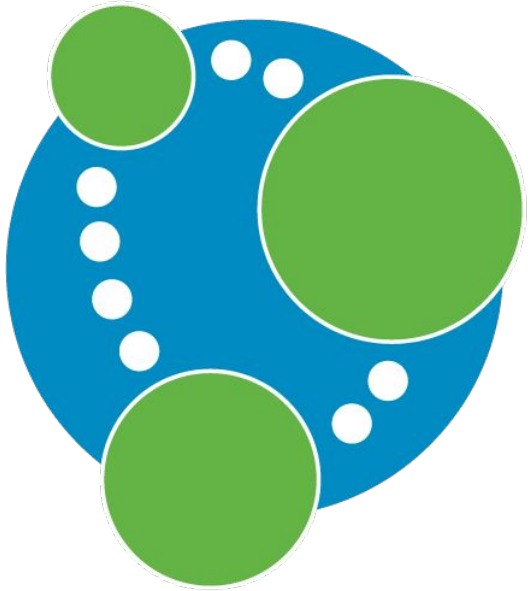
Johannes Unterstein

Software Engineer

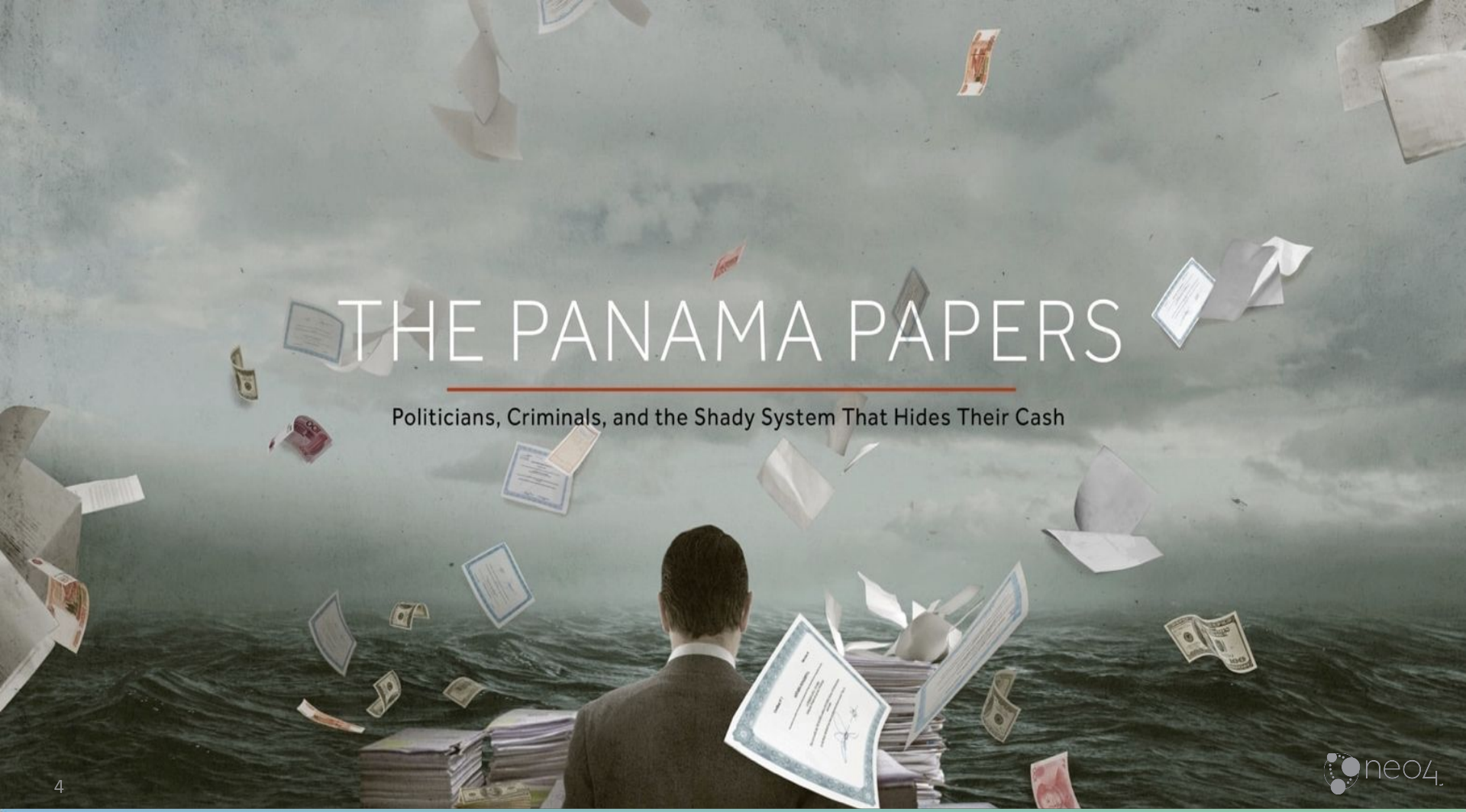
Neo4j.com/Cloud

@unterstein

Neo for what?



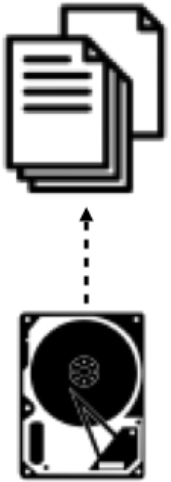
neo4j


THE PANAMA PAPERS
Politicians, Criminals, and the Shady System That Hides Their Cash

Law firm based in panama was leaked 😱

11.5 million documents

Emails, Scanned Documents,
Bank Statements etc…

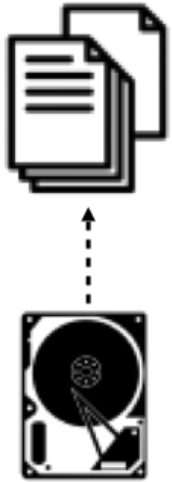


2.6 TB

They tried to analyze all that with excel

11.5 million documents

Emails, Scanned Documents,
Bank Statements etc…

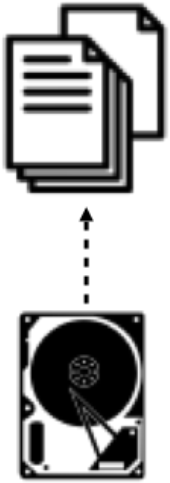


2.6 TB

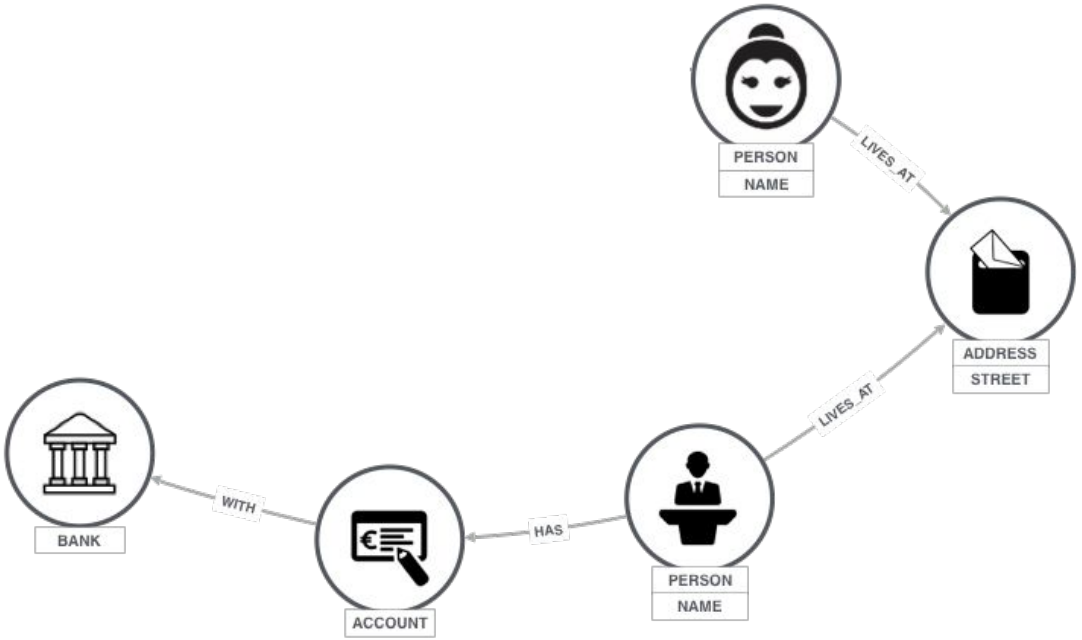
But then they found graphs …

11.5 million documents

Emails, Scanned Documents,
Bank Statements etc…



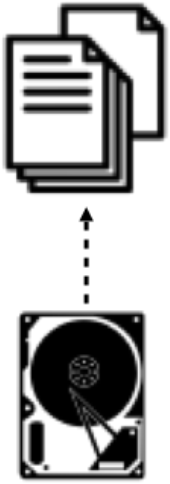
2.6 TB



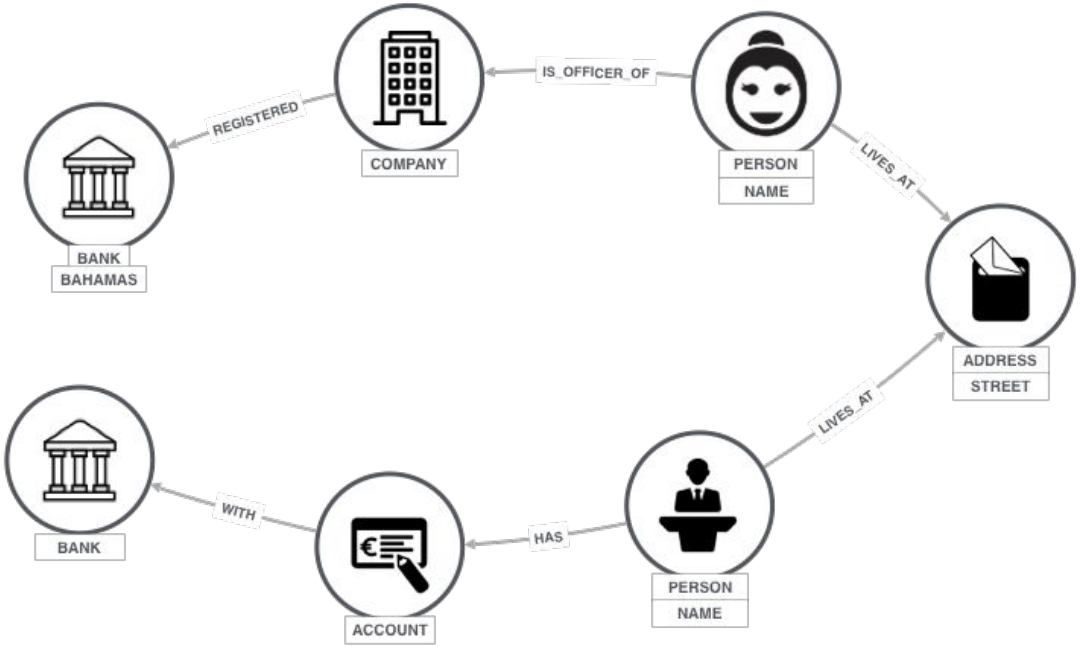
… and the pattern! 😍

11.5 million documents

Emails, Scanned Documents,
Bank Statements etc…



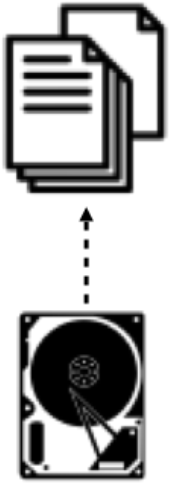
2.6 TB



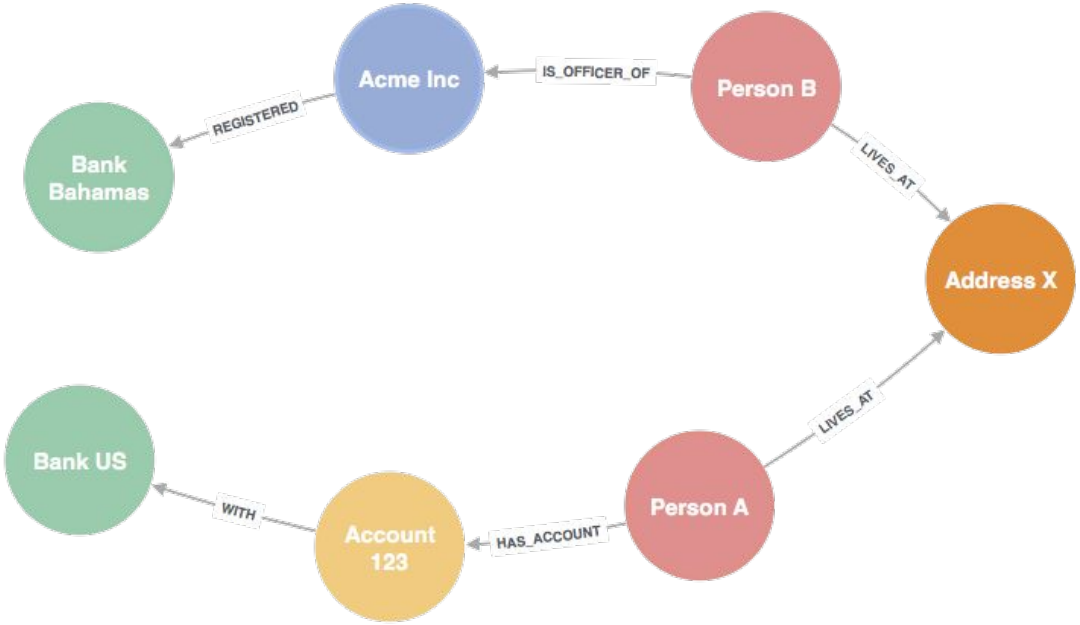
So, this is what we do.

11.5 million documents

Emails, Scanned Documents,
Bank Statements etc…



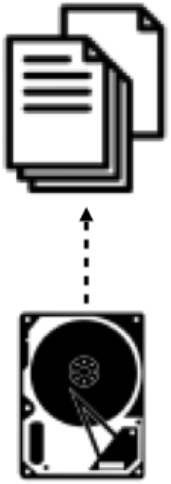
2.6 TB



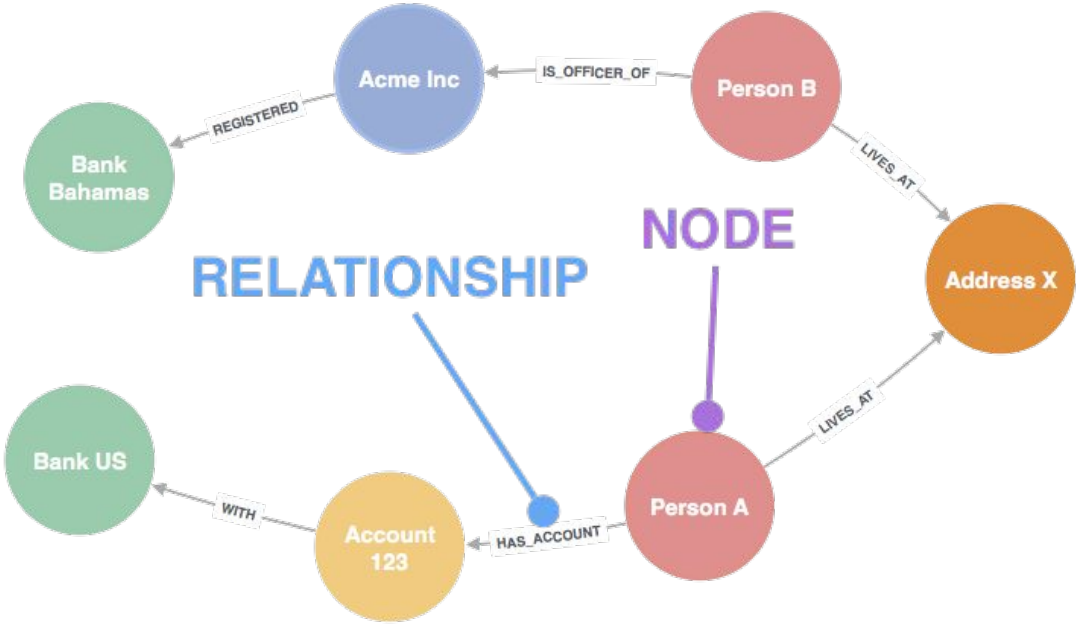
That’s the data model!

11.5 million documents

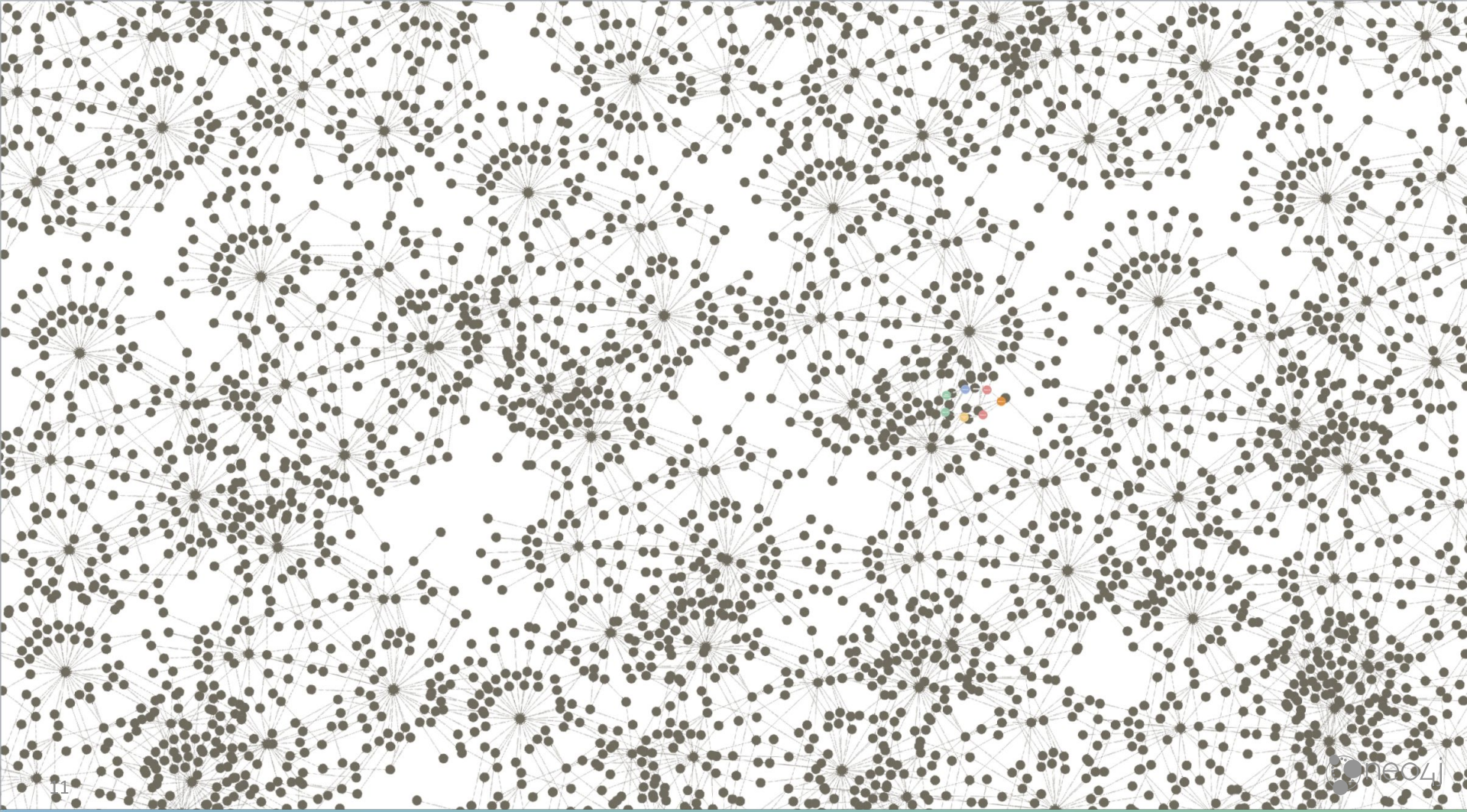
Emails, Scanned Documents,
Bank Statements etc…



2.6 TB



neo4j



Finding a cure for cancer

Johnson Space Center

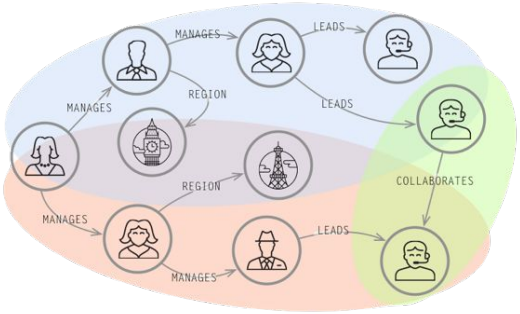
Houston, Texas

Helping to bring people to the mars

More traditional use cases

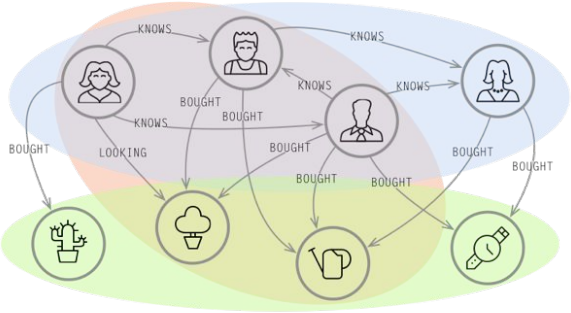
Internal Applications

- Master Data Management
- Privacy, Risk and Compliance
- Fraud Detection



Customer-Facing Applications

- Real-Time Recommendations
- Graph-Based Search
- Identity and Access Management



Why do you put a database in the Cloud?

Why do you put a database in the Cloud?

Kelsey Hightower ✓
@kelseyhightower

Replying to @gabidavila

I also prefer a managed database because of all the other benefits such as backups, upgrades, and high availability, none of which Kubernetes can provide for all databases out of the box.

10:07 PM · Mar 25, 2018 · Twitter Web Client

The story so far

Why change it then?

Running clusters on Kubernetes!

HashiConf 2016

Running clusters on Kubernetes!

Kelsey Hightower 
@kelseyhightower

Most people get really excited about running a database inside [Kubernetes].

This is going to make you lose your job — guaranteed.

HashiConf 2016

Kelsey Hightower
@kelseyhightower

Kubernetes has made huge improvements in the ability to run stateful workloads including databases and message queues, but I still prefer not to run them on Kubernetes.

3:04 PM · Feb 13, 2018 · Twitter Web Client

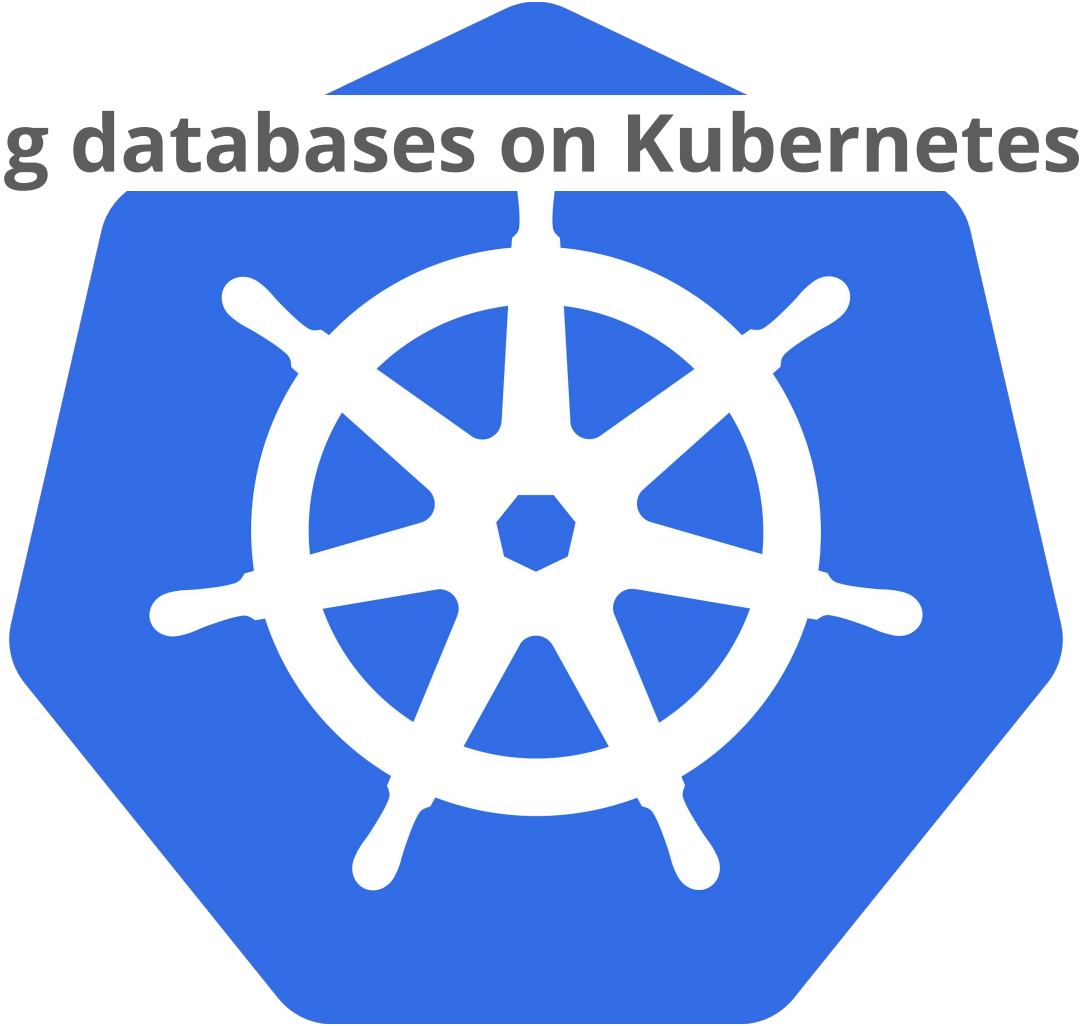
Kelsey Hightower
@kelseyhightower

Replying to @clintkitson

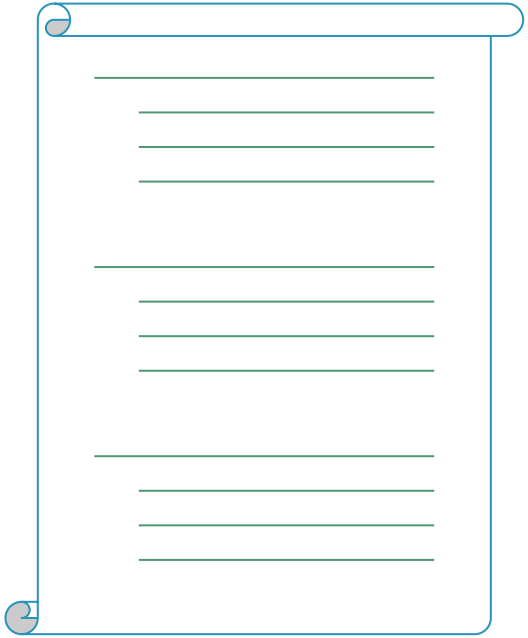
I think it's important to remember that Kubernetes only solves part of the problem. The other parts must be solved by the stateful service and through operational expertise.

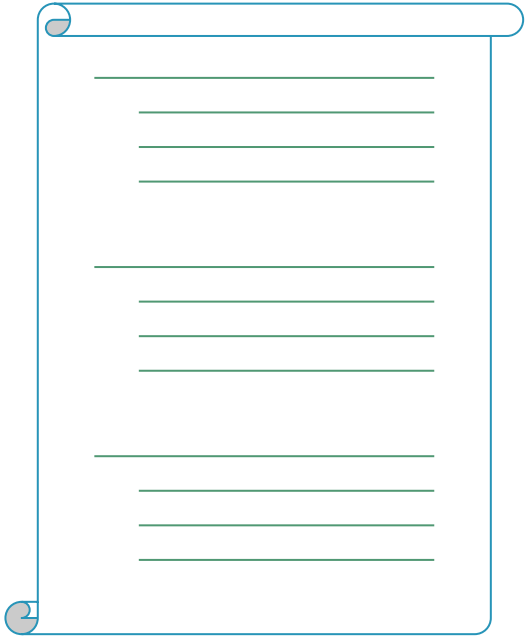
6:09 PM · Feb 13, 2018 · Twitter Web Client

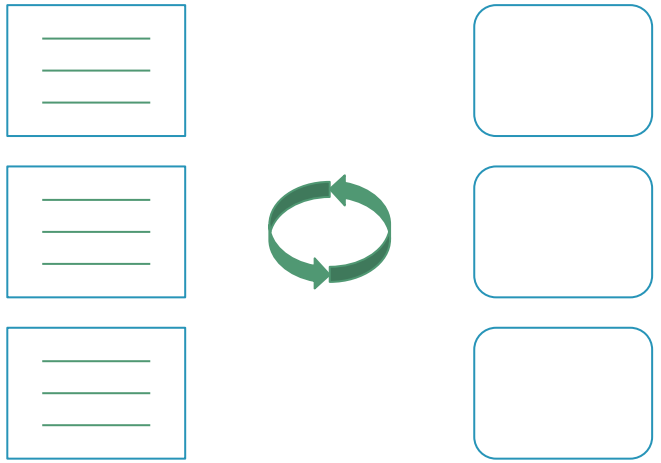
Running databases on Kubernetes!

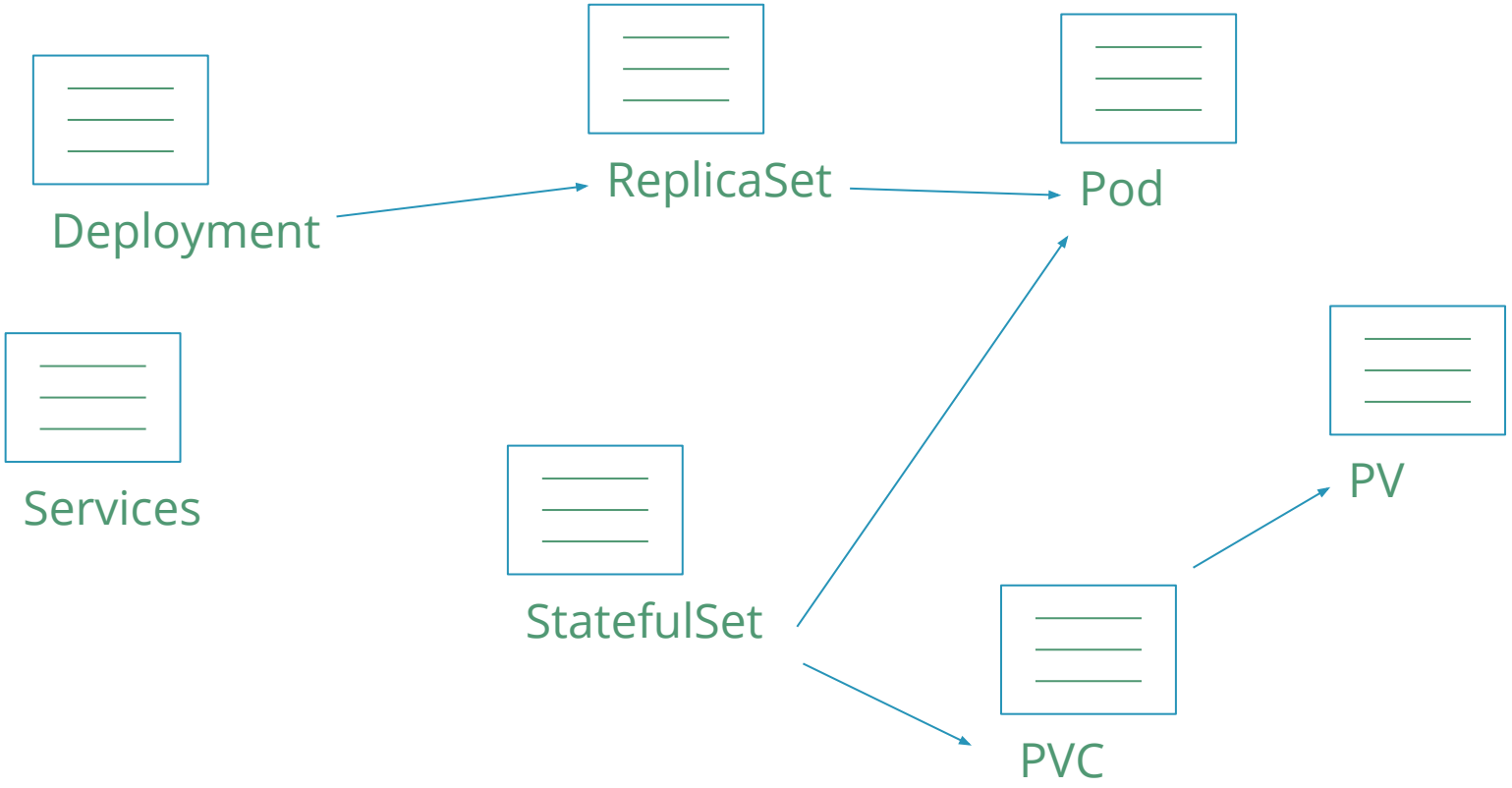


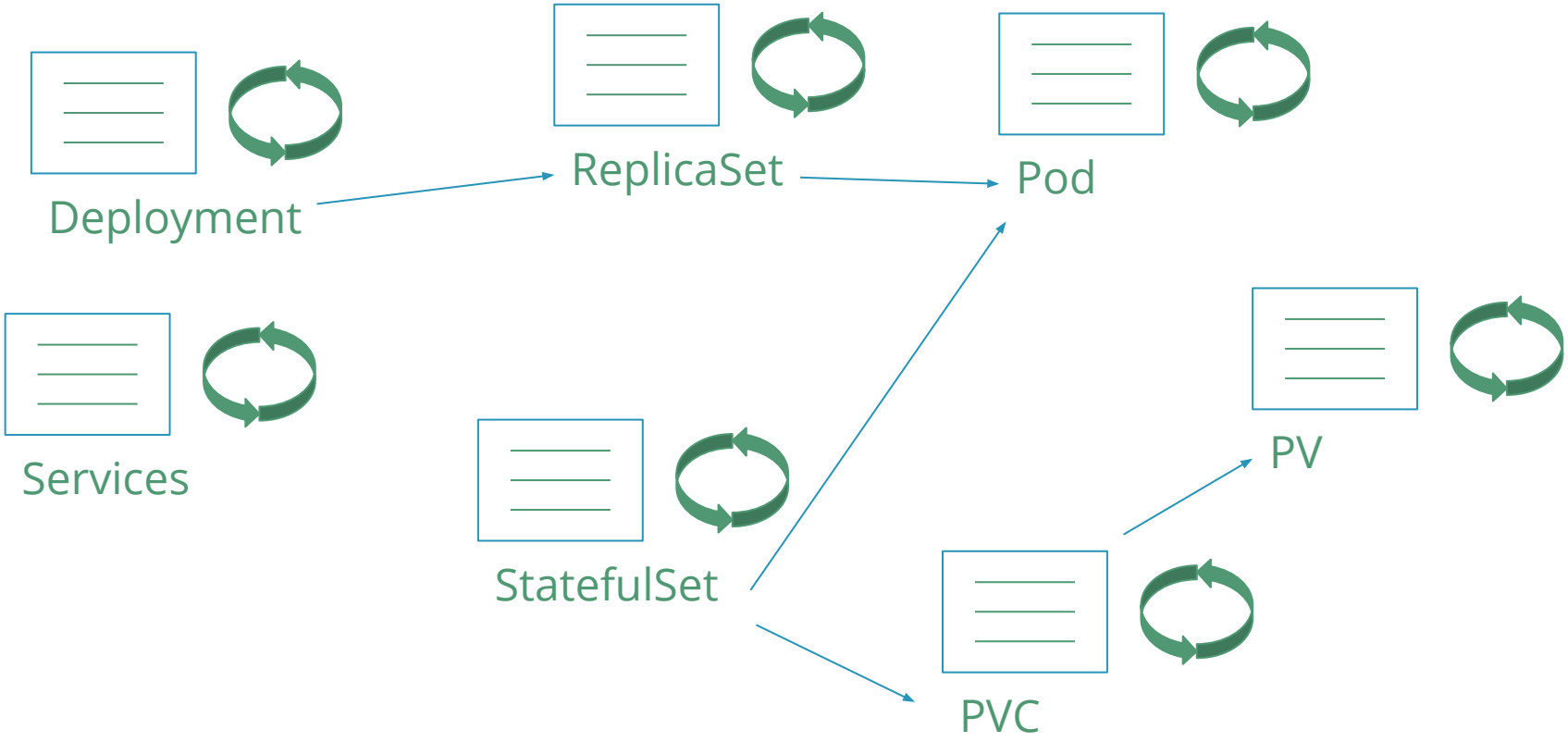
Deep dive Kubernetes

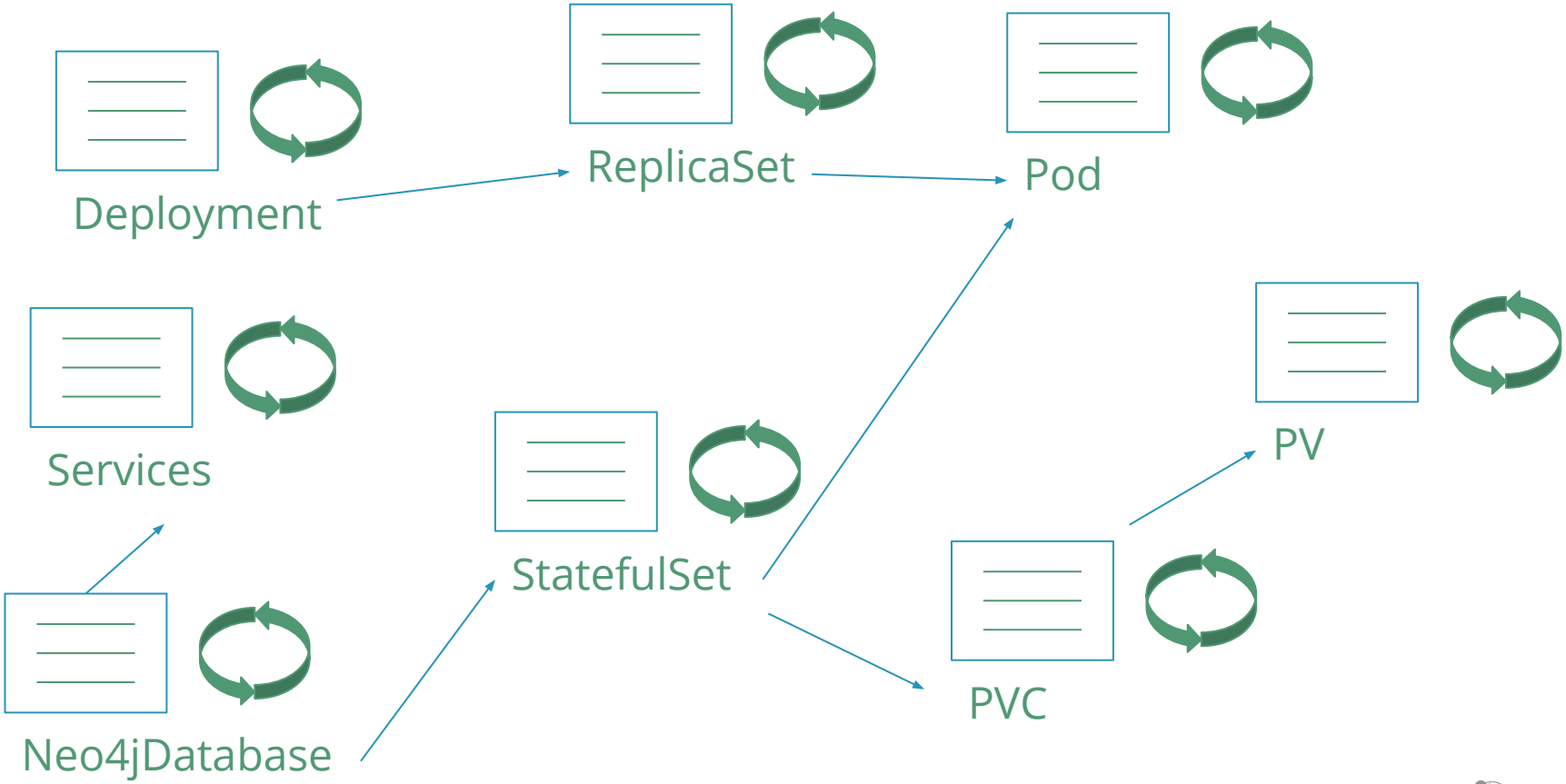






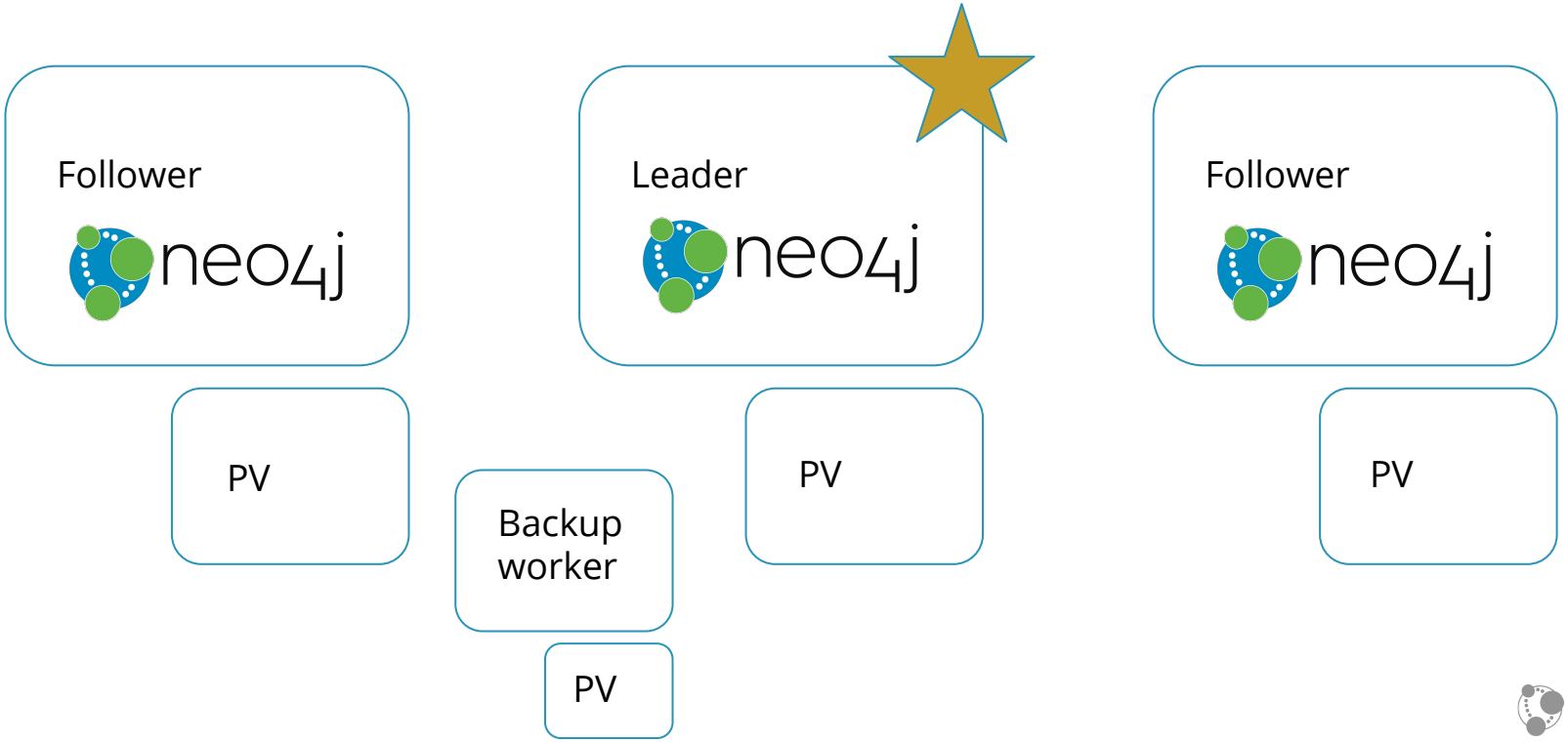



Deployment
ReplicaSet
Pod
Services
StatefulSet
PV
PVC

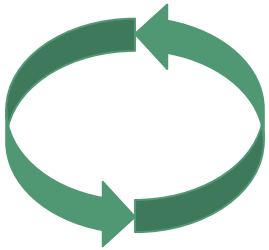

Deployment
ReplicaSet
Pod
Services
StatefulSet
PV
Neo4jDatabase
PVC

The operator

Neo4j clustering



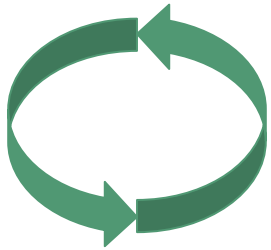
How we structured the reconciler



How we structured the reconciler

backup
load
restore

configs
certs
dns



instances

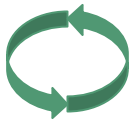
services

The desired state calculation

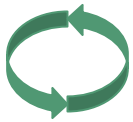
- Main reconciler builds actual current and desired state

- Then it loops over sub reconcilers until:
  - all finished
  - or an error happened
  - or a sub reconciler requests changes

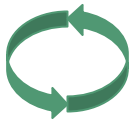
Create



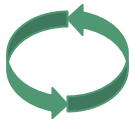
Create



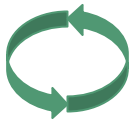
Create



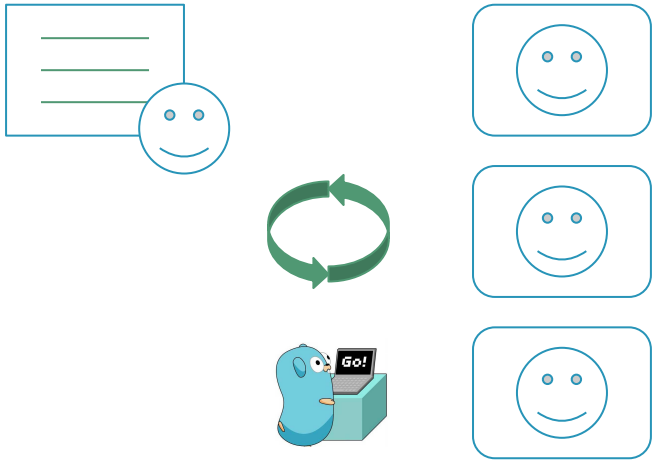
Create



Create

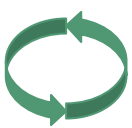


Create

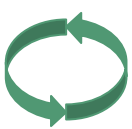


Heal

Heal

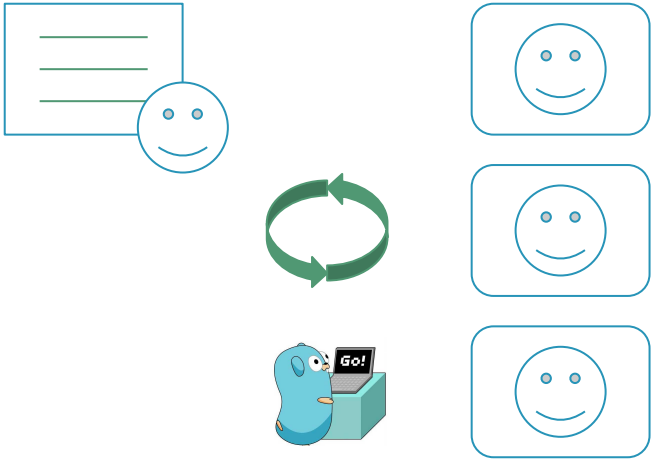


Heal

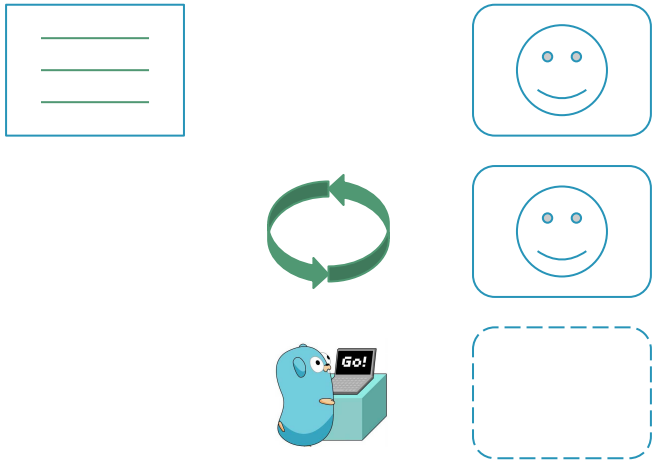


Heal

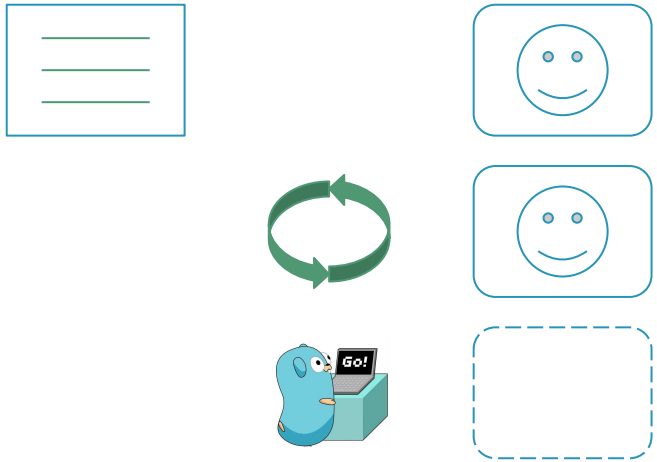
Heal



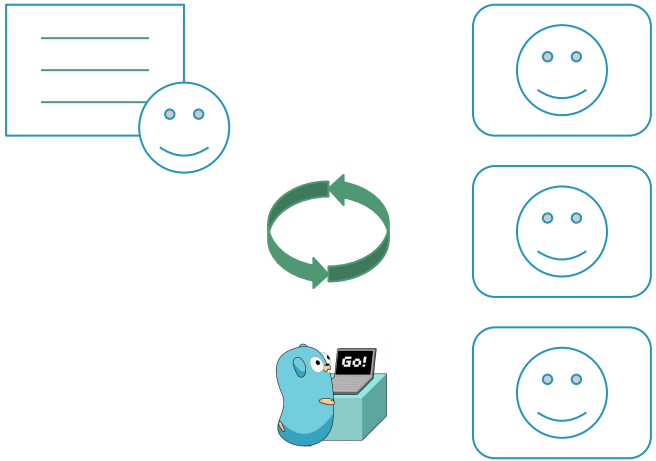
Create



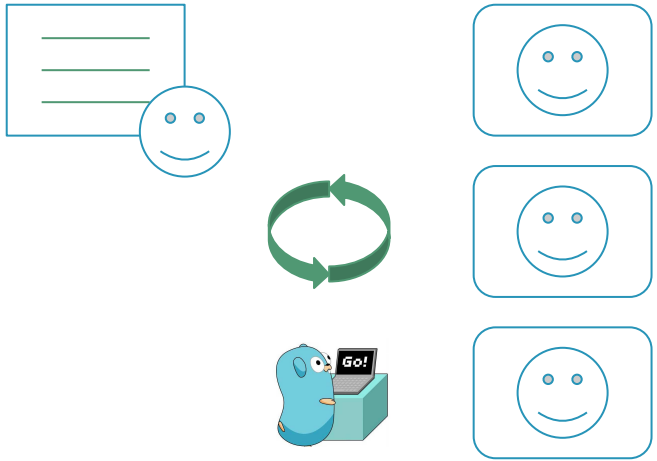
Heal



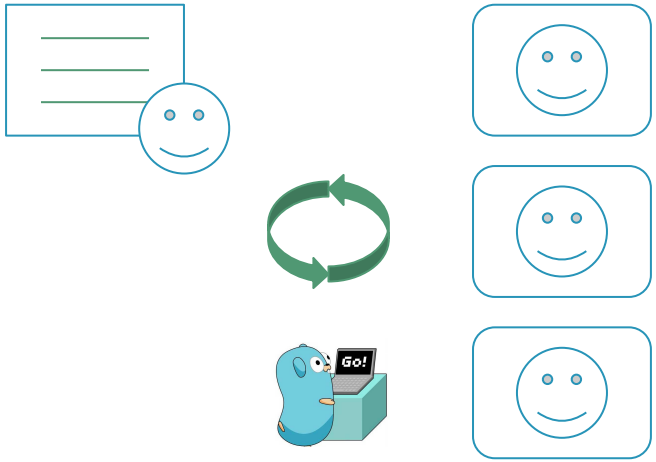
Create



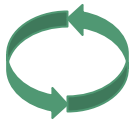
Heal

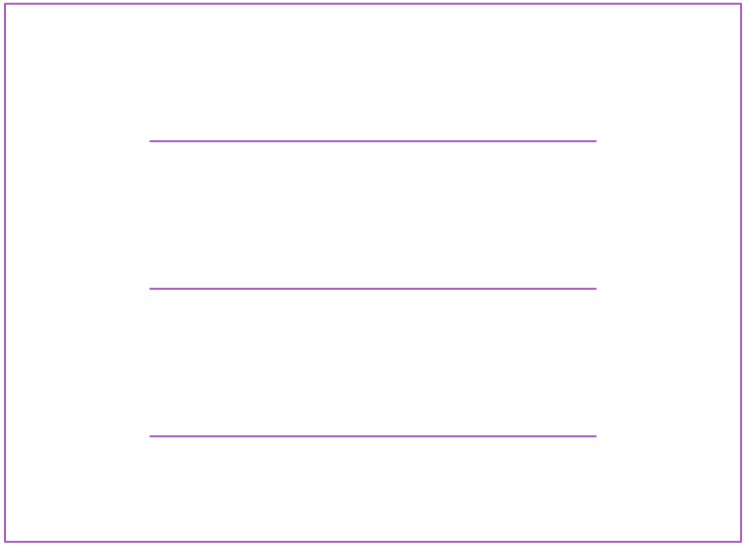


Update

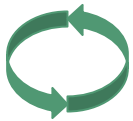


Update





Update

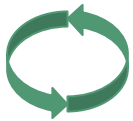


New Neo4j version?

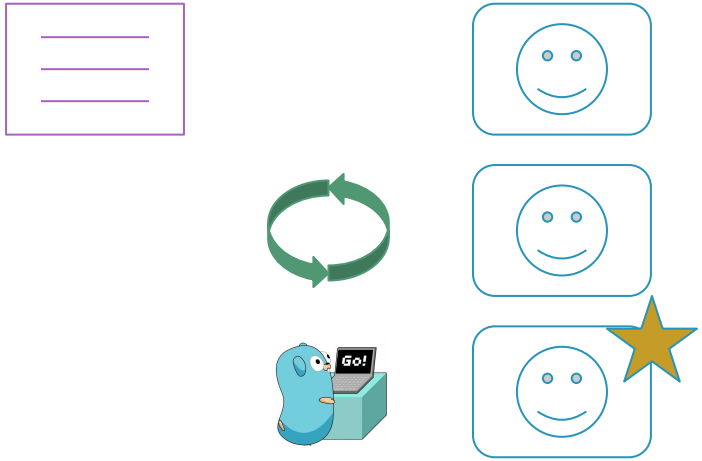
Resource changes?

Password reset?

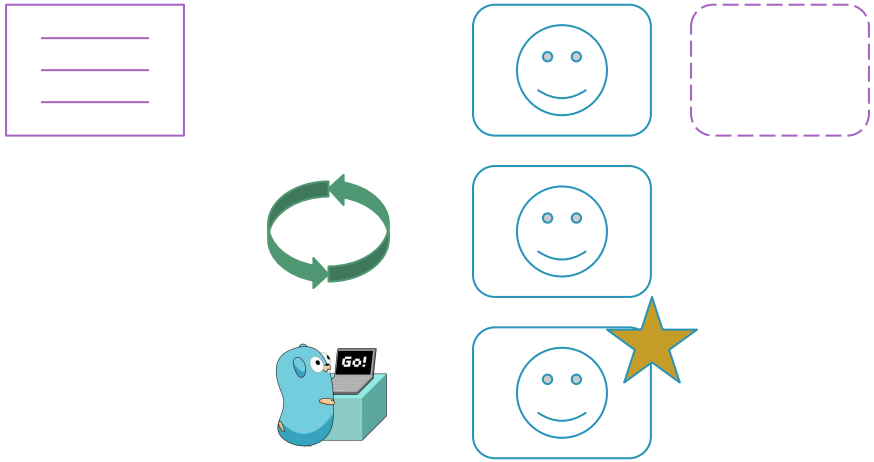
Update



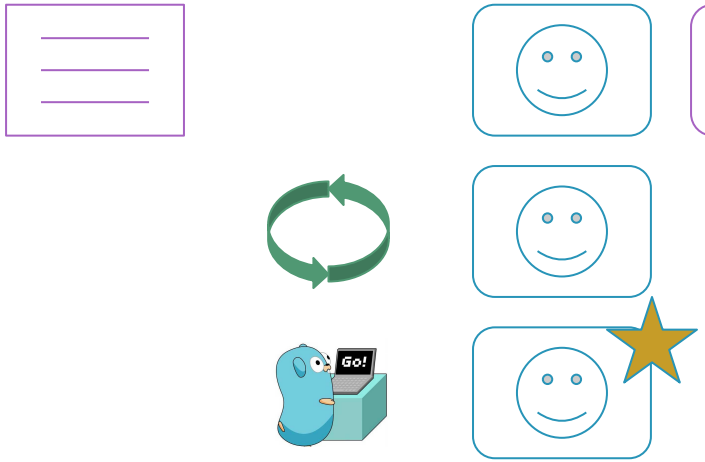
Update



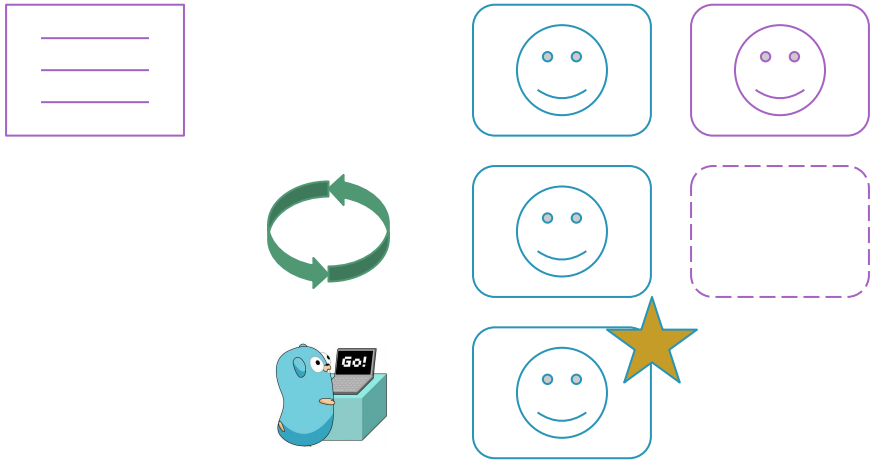
Update



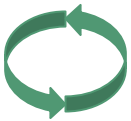
Update

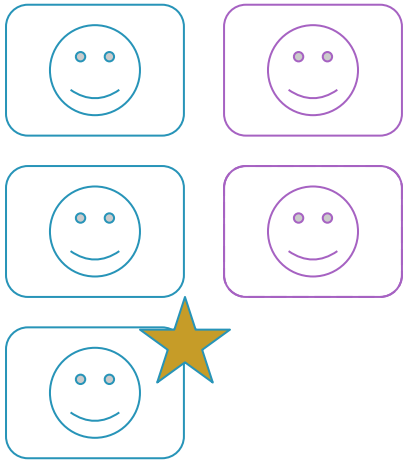


Update

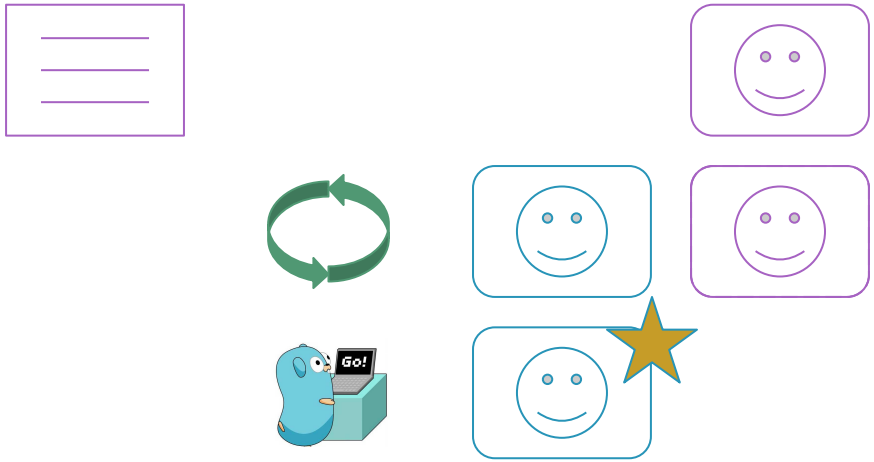


Update

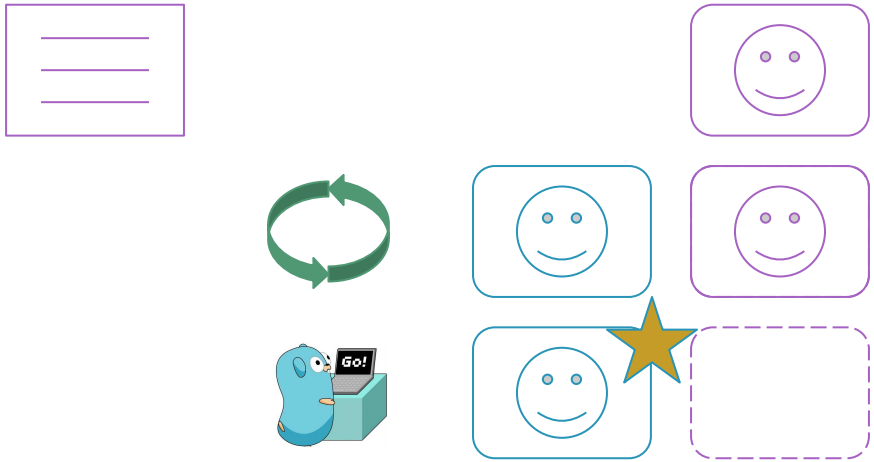




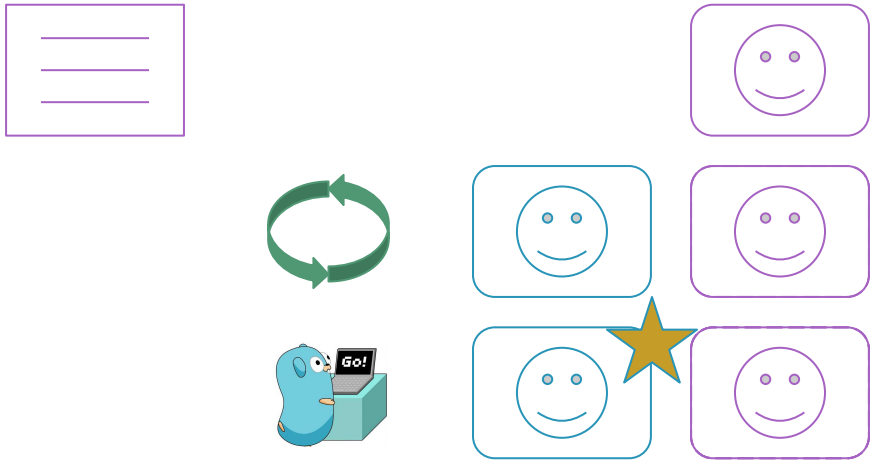
Update



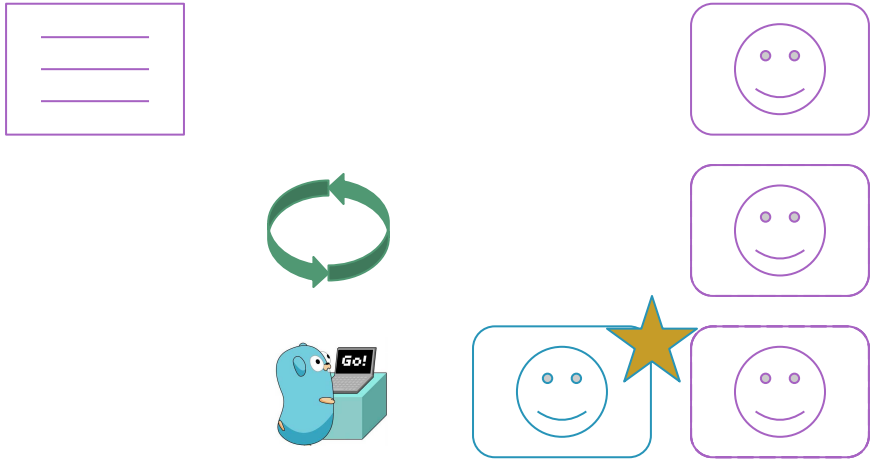
Update



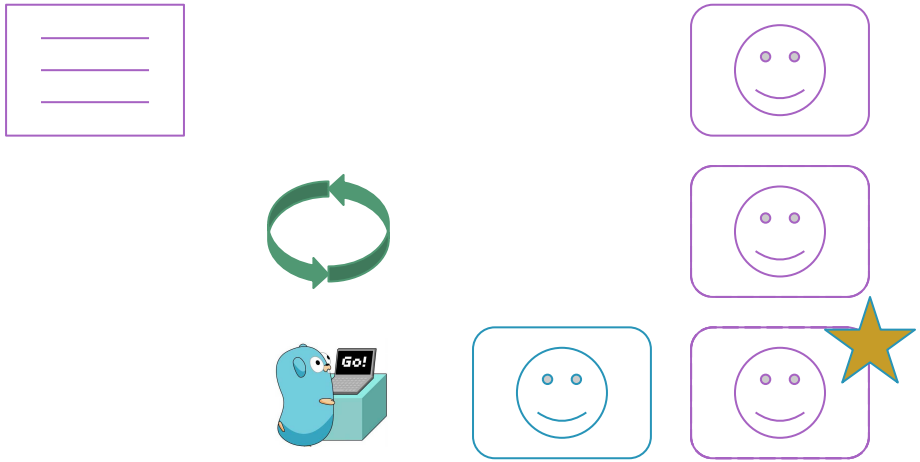
Update



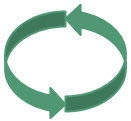
Update

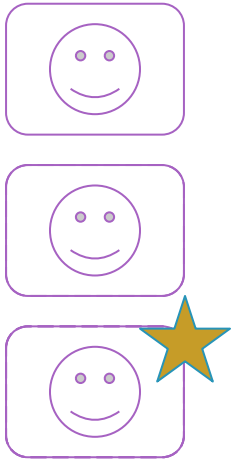


Update

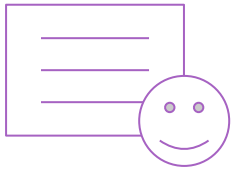


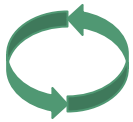
Update

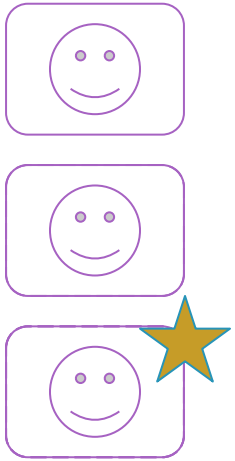




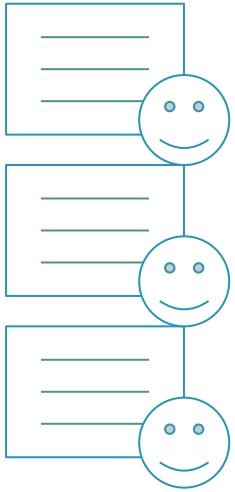
Update

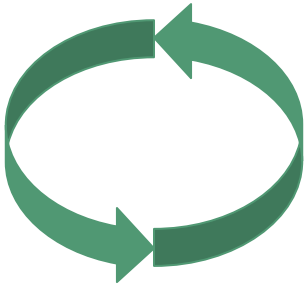


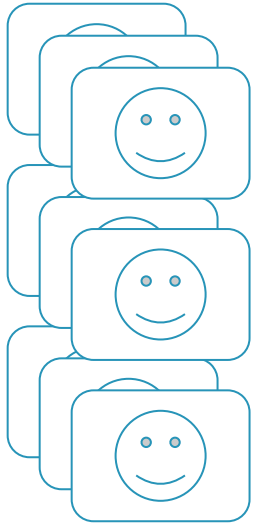




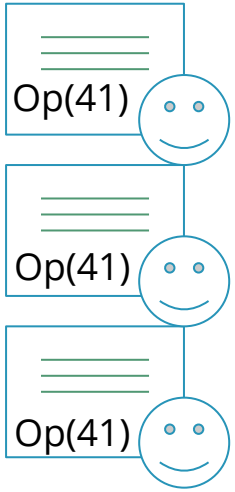
Update the operator itself

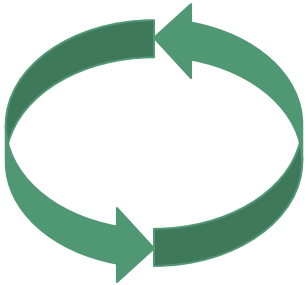




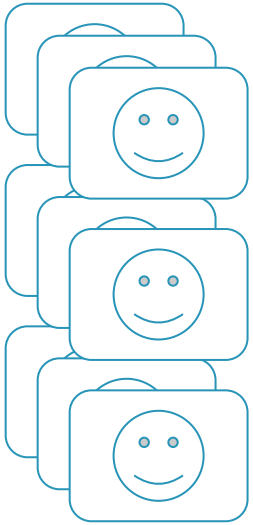


Update the operator itself

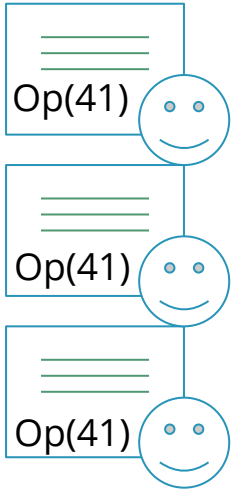


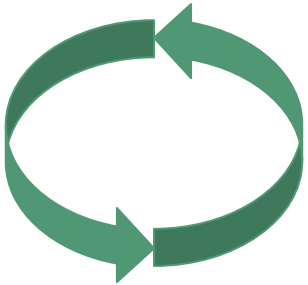


Operator 41



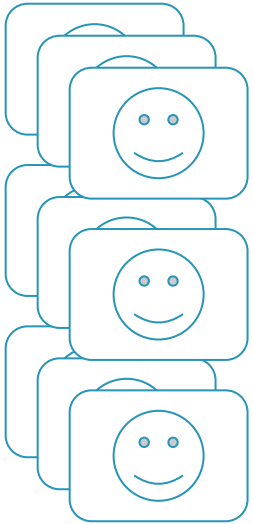
Update the operator itself



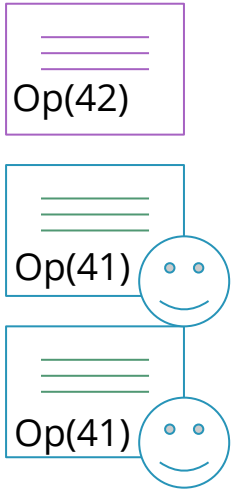


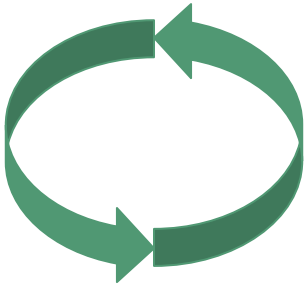
Operator 41

Operator 42



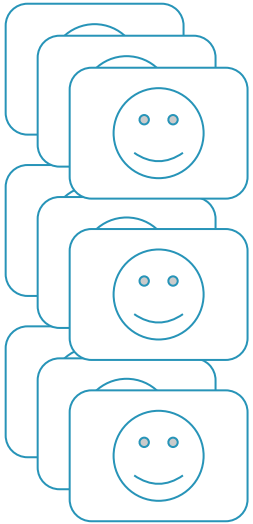
Update the operator itself



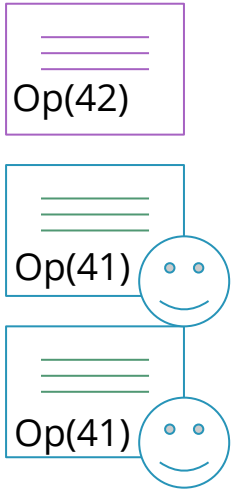


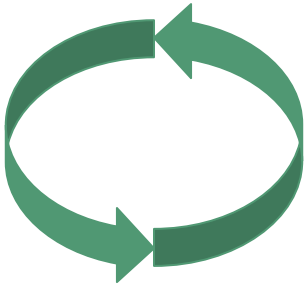
Operator 41

Operator 42



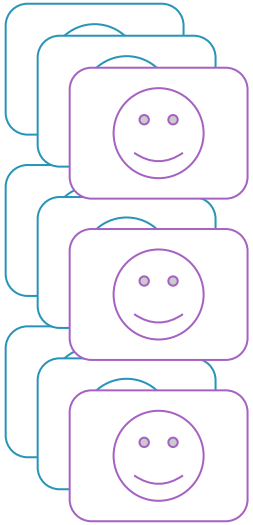
Update the operator itself



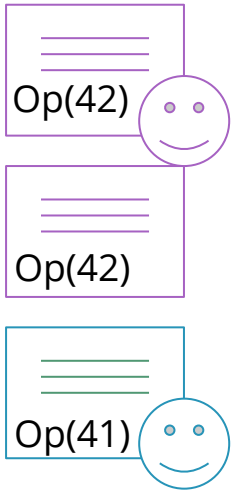


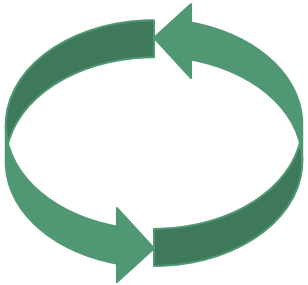
Operator 41

Operator 42



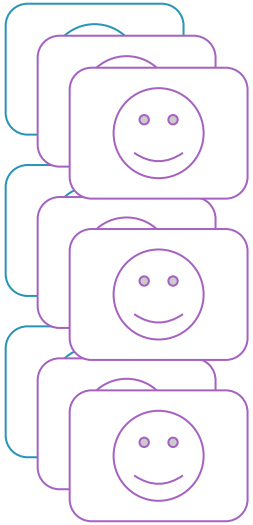
Update the operator itself



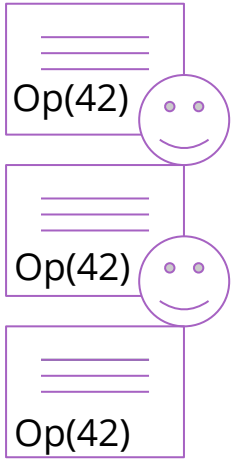


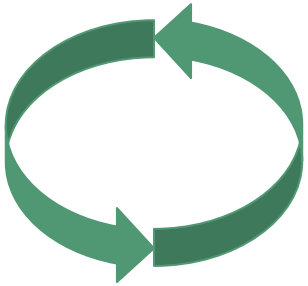
Operator 41

Operator 42



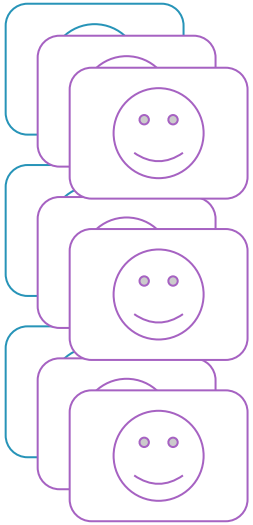
Update the operator itself



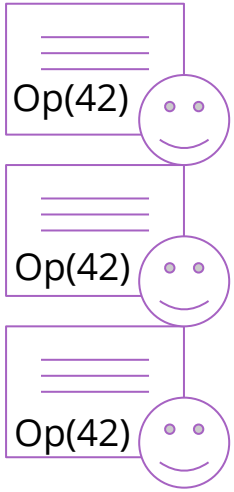


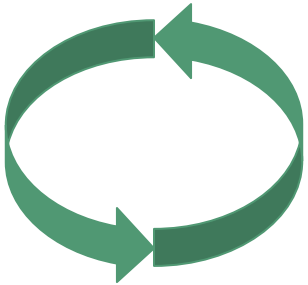
Operator 41

Operator 42



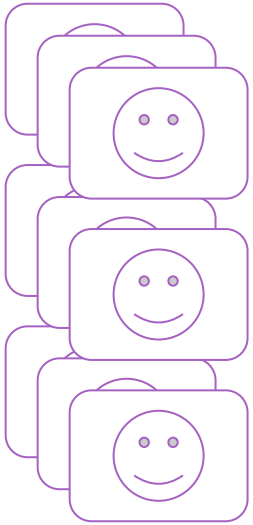
Update the operator itself



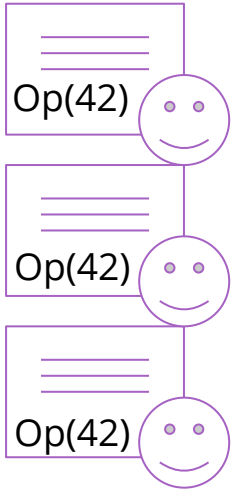


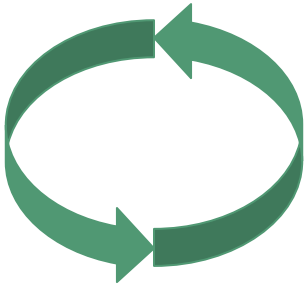
Operator 41

Operator 42

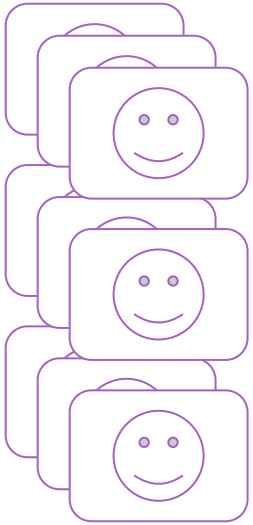


Update the operator itself





Operator 42



So much more to talk about

So much more to talk about

Conditions
Metrics
Monitoring
Alerting

So much more to talk about

Conditions
Metrics
Monitoring
Alerting

Sidecars

So much more to talk about

Conditions
Metrics
Monitoring
Alerting

Sidecars

Debuggability
‘Do not touch’ mode

Is everything happy then?

What about testing?

Unit tests

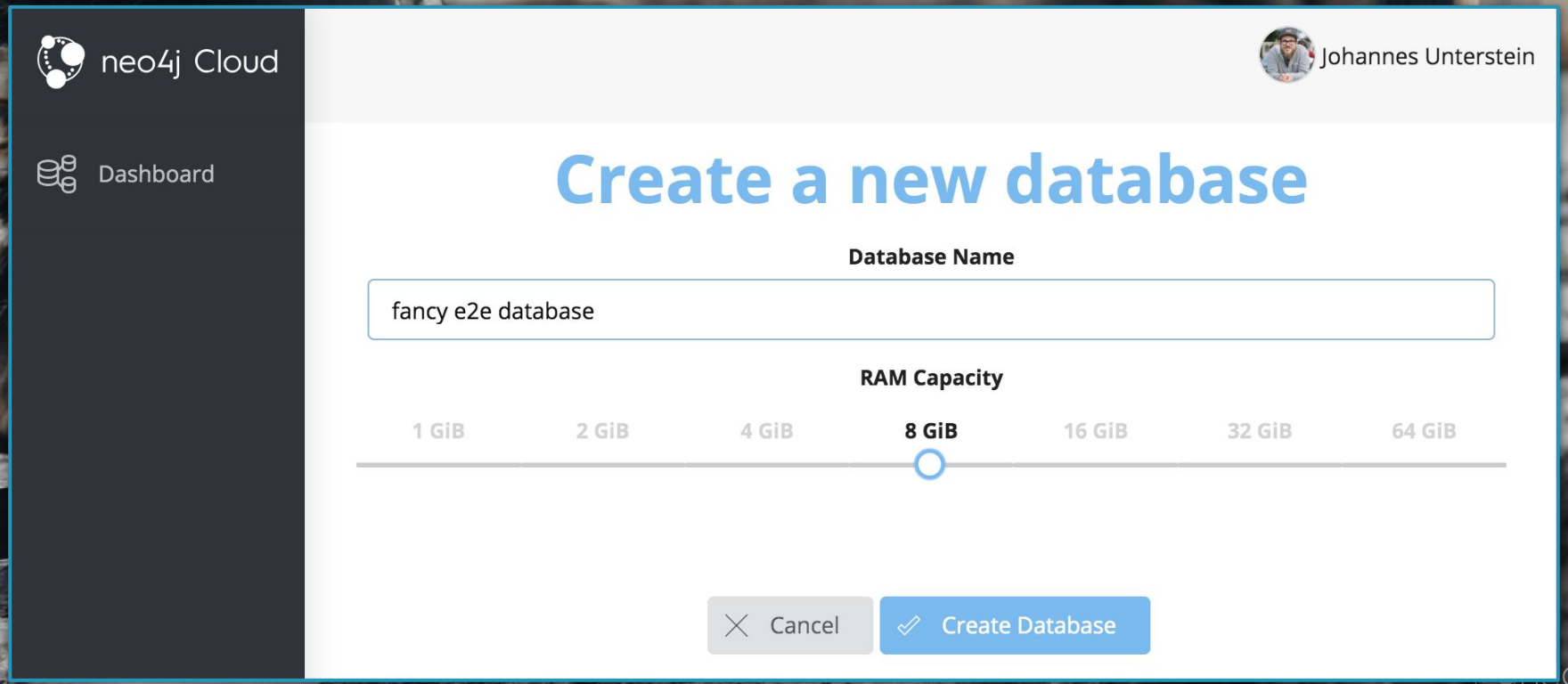
```
tests := map[string]struct {
    states        []*neo4j_status.StatusEndpointResponse
    expected      bool
    boltResponses int
}{
    "expect true when 2 instance have same voting set as leader and 1 has different": {
        states: []*neo4j_status.StatusEndpointResponse{
            healthyResponse(id_2),
            healthyResponse(id_3),
            unhealthyResponse(id_1,  t: 0, id_1),
        },
        expected:      true,
        boltResponses: 2,
    },
    "expect false when 2 leaders": {
        states: []*neo4j_status.StatusEndpointResponse{
            healthyResponse(id_2),
            unhealthyResponse(id_1, typeTwoLeaders),
        },
        expected:      false,
        boltResponses: 2,
    },
}
for name, tc := range tests {
    t.Run(name, func(t *testing.T) {
        actual := AvailableCluster(tc.states, boltResponses, logger)
        assert.Equal(t, tc.expected, actual)
    })
}
```

Integration tests

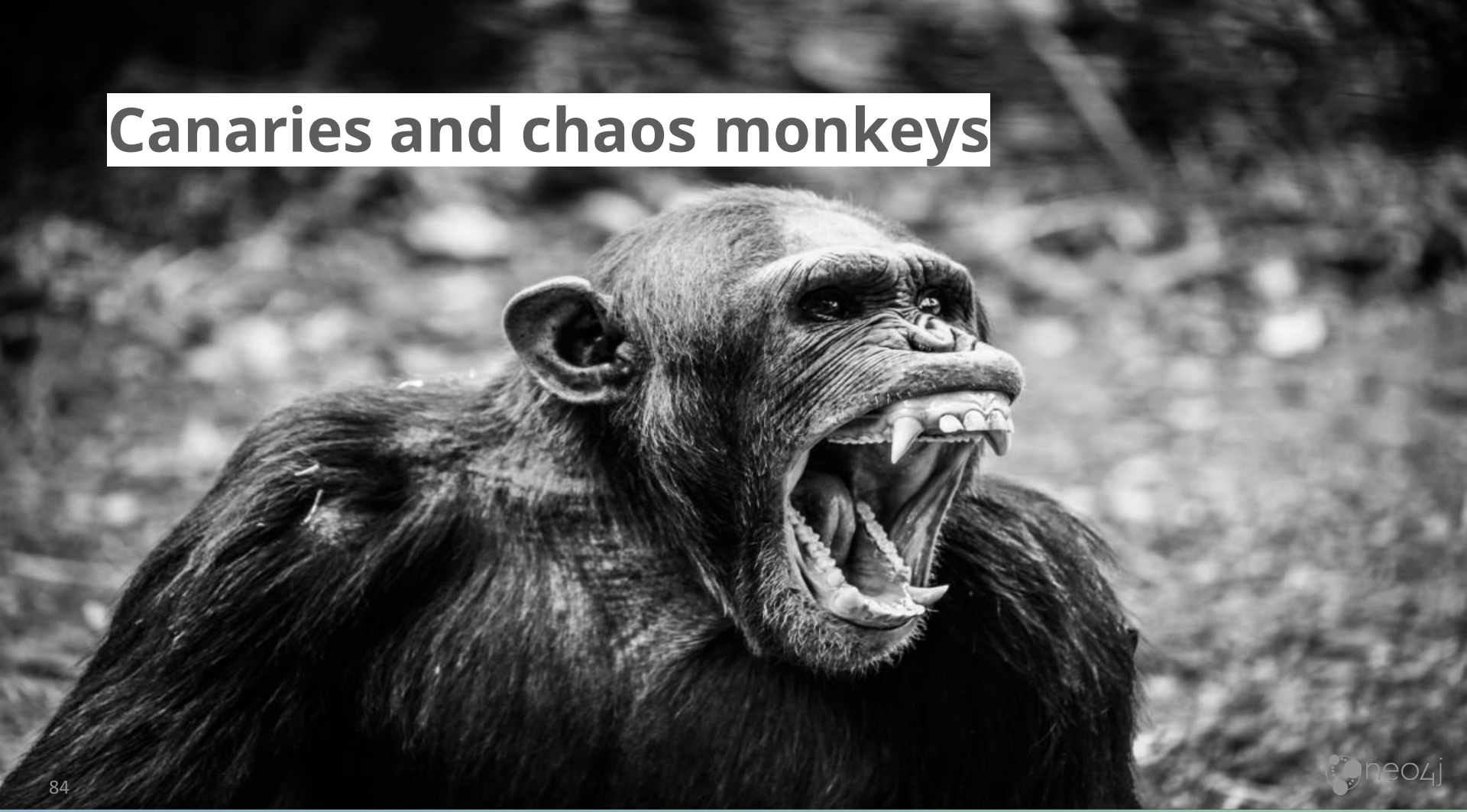
```
func TestCreateConfigMap(t *testing.T) {
    g := setUpConfigMapReconcilerTests(t)

    mockWriteFacade := &context.MockWriteFacade{}

    cluster := newFakeCluster(neo4japi.DBID("cluster-tcmu"))
    neo4jContext := context.NewFakeNeo4jContext(t, mockWriteFacade)

    actual := NoConfigMap
    desired := NewConfigMapState(NewConfigMap(cluster))

    mockWriteFacade.On( methodName: "CreateConfigMap", mock.Anything, mock.Anything).
        Return(noError)

    reconciler := NewConfigMapReconciler(actual, desired, neo4jContext)

    g.Expect(reconciler.Reconcile(cluster)).
        To(BeTrue(), optionalDescription: "Expected reconciliation to be complete")

    mockWriteFacade.AssertCalled(t, methodName: "CreateConfigMap", cluster, desired.ConfigMap())
}
```

System integration tests

Automated E2E tests



Canaries and chaos monkeys



Thank you!

We are hiring SREs and DBaas K8s Devs

👉 l.neo4j.org/dbaas-k8s 👈

@neo4j, @unterstein

neo4j

Pictures from https://pxhere.com 🙏🤘